SIMULATED TESTING OF AN ADAPTIVE MULTIMEDIA INFORMATION RETRIEVAL SYSTEM

Frank Hopfgartner, Jana Urban, Robert Villa and Joemon Jose

Department of Computing Science
University of Glasgow
United Kingdom
{hopfgarf,jana,villar,jj}@dcs.gla.ac.uk

ABSTRACT

The Semantic Gap is considered to be a bottleneck in image and video retrieval. One way to increase the communication between user and system is to take advantage of the user's action with a system, e.g. to infer the relevance or otherwise of a video shot viewed by the user. In this paper we introduce a novel video retrieval system and propose a model of implicit information for interpreting the user's actions with the interface. The assumptions on which this model was created are then analysed in an experiment using simulated users based on relevance judgements to compare results of explicit and implicit retrieval cycles. Our model seems to enhance retrieval results. Results are presented and discussed in the final section.

1. INTRODUCTION

With the improving capabilities of current hardware systems, there are ever growing possibilities to store and manipulate videos in a digital format, leading to a growing number of video archives. People build their own digital libraries from materials created through digital cameras and camcorders, and use systems such as YouTube[1] and Google Video[2] to place this material on the web. Unfortunately, this data creation prowess is not matched by any comparable tools to organise and retrieve video information.

There is a need to create new retrieval engines to assist the user in searching and finding video scenes he/she would like to see from many different video files. Unlike text retrieval systems, retrieval on digital video libraries is facing a serious problem: The Semantic Gap. This is the difference between low-level data representation of videos and the higher level concepts users associate with video.

One way to address this problem is to use the interaction between users and the system. There are different types of interactions, usually divided into two categories: explicit and implicit feedback. Explicit feedback is given when a user informs a system on purpose, what it has to do, such as selecting something and marking it as relevant. Implicit feedback is given unconsciously. An example is printing out a web page, which may indicate an interest in that web page.

In this paper, we present a video retrieval system that utilises explicit *and* implicit relevance feedback from the user. To take advantage of implicit feedback, a model was developed for weighting the different feedback types.

Implicit indicators have been used and analysed in other domains, such as the WWW [1] and text retrieval [2, 3]. In this paper we analyse a model of implicit feedback. In our experiment we look at the advantages the use of implicit feedback can possibly give in retrieval performance, by simulating users based on the collection relevance information. The results were then analysed to investigate the assumptions on which the model of implicit information is based.

This paper is organised as follows: It gives a short survey of existing video retrieval systems in section 2 and presents their inadequacies, which motivated our work. We implemented a video retrieval system which can be used to give explicit and implicit relevance feedback. In section 3, we introduce this system and outline the model of implicit information which is built into the system. Four simulated user studies were performed (section 4) to investigate the type of performance improvement we might hope to gain given the use of the implicit model. In the remainder of this paper, we discuss our simulation results, draw conclusions and focus on our future work.

2. BACKGROUND AND MOTIVATION

Most retrieval systems for digital video libraries are evaluated on the TRECVID test collections, including the Informedia System from Carnegie Mellon University [4], the Físchlár Digital Video System from Dublin City University [5] and the system developed at the Imperial College London [6]. These current approaches are very similar: They use text and visual

[1] http://www.youtube.com/
[2] http://video.google.com/

surrogates to identify results of video shots[3], which are presented by keyframes.

Hauptmann et al. (2004) [8] developed and compared two video retrieval systems using visual *and* textual data versus a visual-only system as part of the Informedia project. In addition, they compared expert and naïve users.

Hauptmann et al. (2005) [9] evaluated the system in low-level feature extraction, semantic concept feature extraction and searching. Their interface visualised a list of results which are associated with text terms. The retrieval results are not dependent on relevance feedback. Their main focus is on comparing the evolution of topics and data set through the years and measuring novice and expert users.

Foley et al. (2005) [10] experimented in automatic and interactive searching. They developed a multi-user system using a DiamondTouch tabletop device. Using the interface, a user can add images as part of the query and select which feature of the image shall be a reference for similar results. They implemented two versions of that system: one with emphasis on efficient searching, the other one on increasing awareness of the users. They conclude that providing awareness cues improves the retrieval performance. However, their interfaces do not support relevance feedback. Search queries have to be refined manually without any automatic reference to former retrieved results.

Heesch et al. (2004) [11] experimented in video retrieval using searching and browsing with an emphasis on user interaction and user navigation. They developed two systems: one including both searching and browsing, the other including searching only. They conclude that adding the browsing functionality increases retrieval performance. The interface of their interactive video library retrieval system unites both visual and textual search queries and the ability of giving relevance feedback. Even though users can give explicit relevance feedback using their system, the knowledge which can be gained from implicit feedback is ignored [11, 6].

Although, in text retrieval, both explicit and implicit relevance feedback techniques are seen as appropriate approach to enhance retrieval results [2, 3].

As far as we know, no group has studied the use of implicit feedback in digital video library retrieval systems. However, traditional issues of implicit feedback can be addressed in video retrieval since digital video libraries facilitate more interaction and are hence amenable to implicit feedback. Previous studies [12] have shown that controlled feedback is more reliable, suggesting that a combination of implicit and explicit feedback might be useful to retrieve effectively. Visual retrieval techniques alone are inadequate and hence, it might be useful to include implicit relevance feedback.

For testing this assumption, we implemented a video retrieval system under the conditions of the TRECVID workshop. Our objective of this work was to understand the role of implicit and explicit features in video retrieval and to find a relevance feedback model which fits best the weighting between explicit and implicit feedback. We assume that the information given implicitly by a user can be used in video retrieval to enhance search results. For testing this, we developed a retrieval system which can make use of implicit relevance feedback. Our second objective was to provide retrieval results handling these implicit features. We developed a first feedback model where we assumed some features as being important for implicit feedback. Based on that, we ran a simulated study to get data for our objective.

[3]A shot is a small fragment of a video, recorded using the same camera and the same angle. Shots can be detected automatically using visual features such as colour, shape and texture [7].

3. SYSTEM DESCRIPTION

We developed a video retrieval system which has two components: A retrieval system back-end and an interface. The retrieval system uses both textual and visual features.

The text consists of the output of an automatic speech recognition software and surrogates such as the date and time of the broadcast, the broadcasting station or the language spoken in the video source (see section 3.3). The Terrier retrieval system [13] is used for indexing and retrieving based on textual components. The BM25 retrieval model was used.

To support visual queries, we extracted several visual features from the provided keyframes of the data set [14]:

- colour layouts
- content shapes
- dominant colours
- edge histograms
- homogeneous textures

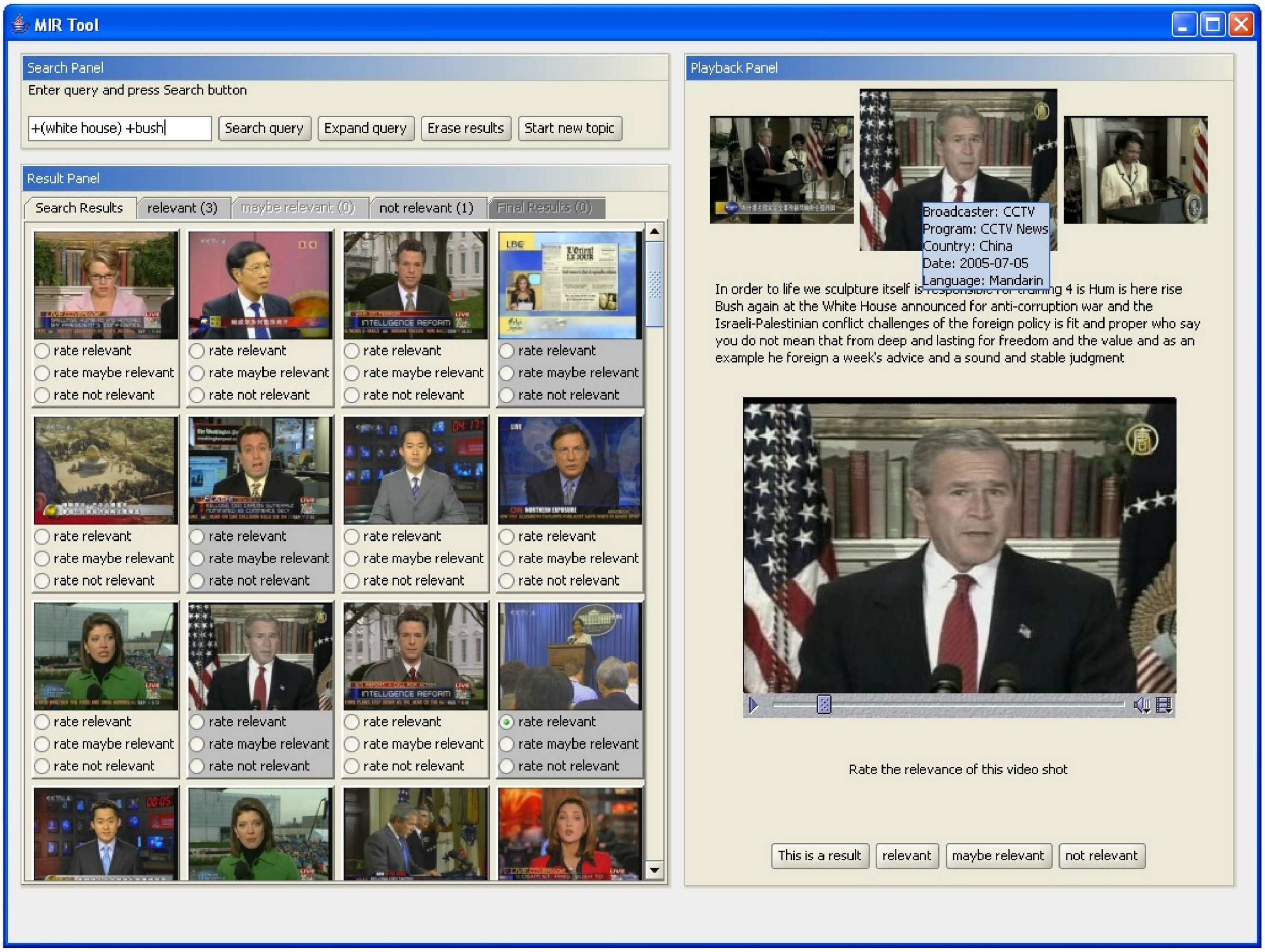
Our retrieval model is based on [15], but adapted according to [16]. We used the voting approach for the combination of evidence from visual and textual features [15].

3.1. Graphical User Interface

As explained before, our objective is to understand the role of implicit and explicit features in video retrieval. Figure 1 shows a screenshot of our interface developed for this purpose.

The interface provides a field for entering a textual query. The query syntax allows use of boolean operators. In response of triggering a query, a list of retrieved shots is presented in the results panel, where a shot is represented by a keyframe. By clicking on a keyframe, the user can play the corresponding video shot and see additional information such as the transcript of the spoken text or video metadata. Keyframes which have been retrieved before will be displayed with another background colour for a better visual identification. A user can give explicit relevance feedback by rating a

Fig. 1. Graphical User Interface



keyframe using the buttons under the keyframes or beneath the video player, respectively.

Three different rating categories are supported: *relevant*, *maybe relevant* and *not relevant*. They are positioned in the result panel as tabs. As a result, retrieval results are grouped into these categories. This grouping has different functions: The *relevant* rated results are used for query expansion. When a user clicks on the *Query Expansion* button, the system suggests query terms that can be added to the initial query. The terms are taken from the video surrogate of the relevant rated keyframes or – if no keyframes have been rated before – from the top 100 results of the initial query (*pseudo relevance feedback*). The user can provide explicit relevance feedback in order to improve retrieval results. The *maybe relevant* group can be seen as a buffer. The user can rate results into this category, if he/she is not sure yet whether this is a relevant result or not. Or he/she can just store results for a later use. The *not relevant* group can be used as a filter. Results which are rated into this category will not appear again in a further retrieval step. This is useful for filtering unwanted results.

Keyframes which have been rated relevant can be used as a visual query in the next retrieval cycle. They can be selected by clicking on the *Query Expansion* button.

In addition to this, we incorporated implicit feedback into the system. This is explained in the next section.

When the user decides to play a video shot, the video and its surrogates will be displayed in the playback panel, which is placed on the right-hand side of the user interface. On the top, a user sees the selected keyframe in context, with its neighboured keyframes to the left-hand and the right-hand side. He can obtain additional information about the video (broadcaster, programme, country, date and language) by moving the mouse over the keyframe. When clicking on the neighboured keyframe, the playback panel will be updated, displaying the video shot and the additional information. Underneath these keyframes, the interface displays the automatic speech recognition text of the selected video shot. Here, the user can mark text and add it to the original search query.

In the middle of the panel, the video shot is played. When the shot reaches its ending time point, the video pauses. The user can start and pause the video at any time by clicking on the play icon under the video. The current playing position is presented by a slider bar. The user can use this bar to navigate in the video file. Furthermore, the user can change the volume and read the Media Properties on clicking on the corresponding icons. Then, a new window pops up which shows additional information like the name of the video file, the duration and the current position.

On the bottom, the user can either mark a shot as a result or rate the relevance of the shot via buttons.

3.2. Mining Interaction Data

As mentioned before, one objective of our work was to develop a system which can make use of implicit feedback. The system monitors user interactions and recommends terms by mining the interaction data. We grouped the interface actions related to the implicit relevance feedback into three categories based on their similarity:

- C_1: Clicking a keyframe in the result list (see the Results Panel in the bottom left corner in figure 1)
- C_2: Playing a video for a specific time slot
- C_3: Interaction with the video (e.g. using the slider bar)

We based these assumptions on [1], assuming that these are some of the most important feedback categories a user can give. Using our interface presented in figure 1, a user *must* click on a keyframe (C_1) to play a video. In doing so, we assume that the user shows interest in the content of that particular part. We also assume that the playing duration is a valid factor to imply interest in a video. The playing duration is divided into specific time slots (C_2), e.g. 5 seconds for each slot. The longer a video is played, the more relevant its content should be. The interaction with the video (C_3) such as using a slider to scroll through it or pressing the pause/stop button shows that the user concentrates on analysing the content of it. We assume that this is another important factor for relevance judgement.

These different features are weighted for measuring the importance of a shot. If more actions appear on the same shot, the weighting should grow, as the implicit factor grows as well. Hence we assume that a shot is more important when a user showed a higher interest in it based on the three defined categories. Some feedback categories appear more often in a user interaction than others, as e.g. playing a video for 10 seconds has the same meaning as playing a video twice for 5 seconds (which is the definition of C_2). Also, 5 interactions with a video would mean five times one interaction with it, which is the definition of C_3. So, the weighting can increase, e.g. depending on the time a video is played.

Basic feedback information, such as "click on a keyframe" or "looking at the metadata", cover a low weighting span.

Giving explicit feedback, a user directly indicates whether a shot is relevant or not. Hence, explicit feedback is more reliable than implicit feedback and therefore should have a higher weighting in our model. As the user might give the implicit relevant feedback unconsciously, it has to be carefully processed to make the correct inferences. Accordingly, implicitly detected results may not receive a higher weighting than explicitly selected. Hence in our model, the contribution of implicit feedback can be combined to a value of 1.0. We define 1.0 as a maximum weighting for explicit relevance feedback. The implicit feedback is aggregated in a strictly monotonic increasing function with values between 0.0 and 1.0. The function we used to achieve this is

$$f(x) = 1 - \frac{1}{x}, where\{x \in \mathbb{R} | x \geq 1\}, \tag{1}$$

and x is the combined implicit relevance feedback weighting a user gave. So, x is a weight resulting from a chain of feedback a user gave implicitly. It can be a possible combination of all feedback categories, e.g. (C_1) + (C_2) + (C_2) + (C_3). In this example, a user would have clicked on a keyframe, played it for two time periods and interacted with it once using the sliding bar.

The following will explain the model using another example: A searcher uses the interface for retrieval in a digital video library. For several results, he/she gives an explicit relevance feedback. These results receive an explicit weighting of 1.0. The searcher does not give any more explicit feedback, but interacts with more results. This interaction is implicit relevance feedback. Each action of these three categories of our model has an implicit relevance feedback weighting (an implementation example is given in section 4). Let's say, every action has a weighting of 1. So, a result the user clicked on (C_1) and played for a specific time (C_2) will have a combined implicit relevance feedback weighting x of 2 (1 for action C_1 plus 1 for action C_2). Using our before mentioned formula, this result will have an implicit weighting of

$$f(2) = 1 - \frac{1}{2} = 0.5$$

The more implicit feedback, the higher $f(x)$, e.g.

$$f(4) = 0.75$$

3.3. Test Collection

Our work was built using the 2005 TRECVID data set and experimented on the TRECVID data set from 2005 and 2006. The 2006 set is approx. 160 hours of television news from November 2004 in English, Chinese and Arabic language, amounting to approx. 130 GB of video data. The data set also includes the output of an automatic speech recognition system, the output of a machine translation system (Chinese and Arabic to English) and the master shot reference. A common set of keyframes is also included.

Each shot is considered as a separate document and is represented by text from the speech transcript. Some statistics:

- 79484 number of shots
- 15.89 terms on average per shot
- 31583 empty shots (without annotation)

The collection also contains search topics and relevance judgements, designed to represent different types of queries real users pose: request for video with specific types of people, specific instances of objects, specific activities or locations [17]. The queries are always in imperative form, examples are presented in table 1:

Table 1. Example search topics of the TRECVID 2006 data set

Find shots of US Vice President Dick Cheney.
Find shots of multiple people in uniform and in formation.
Find shots of US President George W. Bush, Jr. walking
Find shots of one or more people reading a newspaper.
Find shots of something burning with flames visible.
Find shots of a greeting by at least one kiss on the cheek.
Find shots of Condoleeza Rice.

4. SIMULATED EXPERIMENTS

4.1. Experimental Approach

The aim of our work is to provide retrieval results handling implicit features for relevance feedback. In addition, we want to define a model that represents the weighting factors for the different implicit feature categories we introduced in section 3.2. In order to develop a retrieval method, we employed a simulated evaluation methodology which simulated users giving implicit relevance feedback. Therefore, we implemented four different systems S_1–S_4 – one providing explicit relevance feedback (S_1), the other three (S_2–S_4) providing implicit relevance feedback which we classified as introduced in section 3.2. All implemented systems had the same interface. For testing our hypothesis on the developed systems, four different test runs have been carried out. In each test run, a searcher is simulated using the system to perform retrieval with each of the 24 TRECVID topics from either 2005 or 2006.

An initial textual query was given to the retrieval engine based on the search topic and after this first retrieval, the top five relevant results were taken for automatic query expansion. (Relevant shots were detected by comparing the retrieval results with ground truth data.) The idea behind this is that a user would click only on those results which appear to be relevant. The retrieval is then started again with an updated query (with a maximum of six terms – the top six terms that were detected so far) and again, the top five new results which have not been considered before are used as source for a query expansion. These steps were repeated up to 10 times.

In the systems S_2–S_4, we simulated different user behaviour on the top five new results. A user behaviour is divided into different actions, each action is associated with a weight (see table 2), which are used to determine the overall term weights of the shots' index terms. So, the top five results receive a different weighting. In the system S_1, no weighting was given for the results. Our experimental approach is oriented on [2].

Table 2. Weighting of implicit features

Action	W (S_2)	W (S_3)	W (S_4)
Click (C_1)	1	1	10
Playing (C_2)	5	1	5
Interaction (C_3)	10	1	1

We defined *1* as the minimal single feedback weight and *10* as the highest single feedback that can be given with one interaction. In using these weights for the different categories, we receive a broad quantity of normalised weighting factors.

User behaviour was modelled by combining actions from categories C_1, C_2 and C_3. In each system, the categories had different weighting in relation to the other categories: S_2 using $C_1 < C_2 < C_3$, S_3 using $C_1 = C_2 = C_3$ and S_4 using $C_1 > C_2 > C_3$. We supported three different user behaviour cases:

- C_1 and $C2$ (likelihood: 50 %)
- C_1, C_2 and C_3 (likelihood: 40 %)
- C_1, C_2, C_3 and C_3 (likelihood: 10 %)

Possible simulated behaviours are e.g. "Click on keyframe" and "Playing a video" (which adds the weighting of 6 (=1+5) to the retrieved terms using S_2) or "Click on keyframe", "Playing a video" and "Interaction with video" (which adds the weighting of 16 (=1+5+10) to the retrieved terms using system S_2). As a refined query consists of the top six weighted terms, the simulated user behaviour influences the new query implicitly. Using formula 1 proposed in section 3.2, a term with the combined weight of 16 has a normalised weighting of:

$$f(16) = 1 - \frac{1}{16} = 0.9375$$

We assume that the higher weighted terms in the simulation can be equated with terms achieved based on explicit feedback.

4.2. Results

Figure 2 illustrates the results of our simulated tests for the 2006 data set. It displays the total number of retrieved relevant shots over all queries over the relevance feedback iterations for the systems S_1–S_4. As illustrated, the systems S_2–S_4 tend to – apart from few deviations – return higher numbers of retrieved relevant shots over all queries than S_1.

Fig. 2. Total number of retrieved relevant shots over all queries

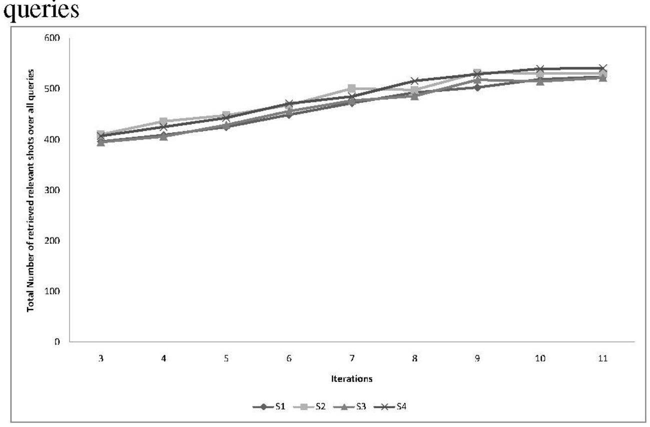
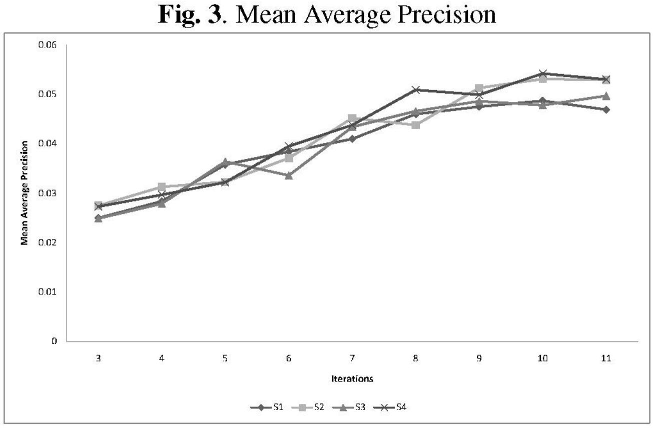


Fig. 3. Mean Average Precision



Our three category model seems to be supported by the mean average precision, as system S_1 (which simulates a user giving explicit feedback only) has the lowest mean average precision (see figure 3). The 2005 data set confirms these results and hence is not shown here. As these figures illustrate, S_4 retrieves a higher number of results than S_1–S_3. So, a model for weighting implicit feedback using our categories should weight $C_1 > C_2 > C_3$.

5. CONCLUSION

We presented two objectives in this work. One objective was to develop a system which can make use of implicit relevance feedback. Therefore, we developed a new video retrieval system and presented a relevance weighting model. The model introduced in section 3.2 was based on research results from textual retrieval. In our model, we classified implicit user interactions into three different categories C_1 (click on a keyframe), C_2 (playing duration) and C_3 (interaction with a video). These categories can be weighted and cumulated, as a user may perform several of these interactions. Their cumulated weighting can express the expanding relevance of a result. The more implicit feedback a user gives on a result, the more relevant it is.

Our second objective was to perform a study using our model. Our idea was to test whether a system using our implicit features returns better retrieval results than a system providing explicit relevance feedback only. An interesting question also was, which weighting should be better for the different categories. Therefore, we ran a simulated user study which was based on our weighting model. In our simulation, a video retrieval system providing both explicit and implicit relevance feedback returned better retrieval results than a system using explicit feedback only. These results support our assumption that implicit relevance feedback may enhance retrieval results.

We emphasised different implicit relevance feedback types in our model. Focusing on these feedback types we conclude that a model giving the initial click on a keyframe C_1 with a higher weighting factor than the view of a keyframe C_2 and the interaction with a video C_3, $C_1 > C_2 > C_3$, retrieves better results than systems using another model.

6. FUTURE WORK

Currently, our assumptions are only supported by our simulated user study. One of the next steps will be to perform a real user study, which may support our assumptions. Only a real study will provide acceptable data to support our objectives. Giving each user a feedback questionnaire will also give useful indices about the acceptance of our interface. It will be interesting to see what they like and what they dislike about it.

Future work will also concentrate on identifying more implicit relevance feedback factors and to implement them into the system.

In addition, further work into the differences in utilising implicit information in a video retrieval interface is required, given the differences between the results found here and those in other domains, such as text retrieval (e.g. [2]). A possible solution we are going to investigate is to exploit ways of interacting with the video more directly (e.g. by tools suggested by [18] and then to use these interactions as implicit feedback factors.

In addition, future work will focus on designing several interface sketches which could back up the findings of our study. Several interfaces could use different implicit feedback factors to show whether our assumptions can be applied to other interface types as well.

7. ACKNOWLEDGEMENTS

This research was supported by the European Commission under the contracts FP6-027026-K-SPACE and FP6-027122-SALERO. It is the view of the authors but not necessarily the view of the communities.

8. REFERENCES

[1] Mark Claypool, Phong Le, Makoto Wased, and David Brown, "Implicit interest indicators," in *Intelligent User Interfaces*, 2001, pp. 33–40.

[2] R.W. White, J.M. Jose, C.J. van Rijsbergen, and I. Ruthven, "A Simulated Study of Implicit Feedback Models," in *Proceedings of the 26th European Conference on Information Retrieval Research (ECIR '04). Lecture Notes in Computer Science*, 2004.

[3] Diane Kelly and Jaime Teevan, "Implicit feedback for inferring user preference: A bibliography," *SIGIR Forum*, vol. 32, no. 2, 2003.

[4] Mike Christel and Ronald Concescu, "Addressing the Challenge of Visual Information Access from Digital Image and Video Libraries," in *Proc. ACM/IEEE-CS Joint Conference on Digital Libraries (Denver, CO, June 2005)*, 2005, pp. 69–78.

[5] Eddie Cooke, Paul Ferguson, Georgina Gaughan, Cathal Gurrin, Gareth J.F. Jones, Herv Le Borgne, Hyowon Lee, Sen Marlow, Kieran Mc Donald, Mike McHugh, Noel Murphy, Noel E. O'Connor, Neil O'Hare, Sandra Rothwell, Alan F. Smeaton, and Peter Wilkins, "TRECVID 2004 Experiments in Dublin City University," in *TREC2004 – Text REtrieval Conference, Gaithersburg, Maryland, 15-19 November 2004*, 2004.

[6] Rui Jesus, J. Magalhães, Alexei Yavlinski, and Stefan Rüger, "Imperial College at TRECVID," in *TRECVid 2005 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 14-15 November 2005*, 2005.

[7] P. Aigrain, H. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review," *Multimedia Tools and Applications*, vol. 3, pp. 179–202, 1996.

[8] Alexander Hauptmann, M.-Y. Chen, Mike Christel, C. Huang, Wei-Hao Lin, T. Ng, Norman Papernick, A. Velivelli, J. Yang, R. Yan, H. Yang, and H.D. Wactlar, "Confounded Expectations: Informedia at TRECVID 2004," in *TREC2004 – Text REtrieval Conference, Gaithersburg, Maryland, 15-19 November 2004*, 2004.

[9] Alexander Hauptmann, Mike Christel, Ronald Concescu, J. Gao, Q. Jin, Wei-Hao Lin, J.-Y. Pan, S.M. Stevens, R. Yan, J. Yang, and Y. Zhang, "CMU Informedia's TRECVID 2005 Skirmishes," in *TRECVid 2005 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 14-15 November 2005*, 2005.

[10] Eddie Foley, Cathal Gurrin, Gareth Jones, Cathal Gurrin, Gareth Jones, Hyowon Lee, Sinad McGivney, Noel E. O'Connor, Sorin Sav, Alan F. Smeaton, and Peter Wilkins, "TRECVid 2005 Experiments at Dublin City University," in *TRECVid 2005 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 14-15 November 2005*, 2005.

[11] Daniel Heesch, Peter Howarth, J. Magalhães, Alexander May, Marcus Pickering, Alexei Yavlinski, and Stefan Rüger, "Video Retrieval using Search and Browsing," in *TREC2004 – Text REtrieval Conference, Gaithersburg, Maryland, 15-19 November 2004*, 2004.

[12] Ryen W. White, J.M. Jose, and I. Ruthven, "Adapting to Evolving Needs: Evaluating a Behaviour-Based Search Interface," in *Proceedings of 17th Annual HCI Conference (2nd Volume) Bath, UK, 2003*, 2003.

[13] Iadh Ounis, Gianni Amati, Vassilis Plachouras, Ben He, Craig Macdonald, and Douglas Johnson, "Terrier Information Retrieval Platform," in *Proceedings of the 27th European Conference on Information Retrieval (ECIR 05), Santiago de Compostela, Spain*, 2005.

[14] Jana Urban, Xavier Hilaire, Frank Hopfgartner, Robert Villa, Joemon Jose, Siripinyo Chantamunee, and Yoshihiko Gotoh, "Glasgow University at TRECVID 2006," in *TRECVID 2006 - Text REtrieval Conference TRECVid Workshop*, 2006.

[15] Jana Urban and Joemon M. Jose, "Evidence combination for multi-point query learning in content-based image retrieval," in *Proc. of the IEEE Sixth Int. Symposium on Multimedia Software Engineering (ISMSE'04)*, Dec. 2004, pp. 583–586.

[16] Z. Xiong, X. Zhou, W. Pottenger, and T. Huang, "Speeding up relevance feedback in image retrieval with triangle inequality algorithm," 2001.

[17] P. Enser and C. Sandom, "Retrieval of Archival Moving Imagery - CBIR Outside the Frame?," in *CIVR '02:*

Proceedings of the International Conference on Image and Video Retrieval, London, UK, 2002, pp. 206–214, Springer-Verlag.

[18] Y. Yamamoto, K. Nakakoji, and T. Akio, "The Landscape of Time-Based Visual Presenation Primitives for Richer Video Experience," in *Proceedings of the International Conference on Human-Computer Interaction*. 2005, pp. 795–808, Springer-Verlag.